Stimulus Equivalence, Generalization, and Contextual Stimulus Control in Verbal Classes

Zuilma Gabriela Sigurðardóttir, University of Iceland
Harry A. Mackay, Shriver Center, University of Massachusetts Medical School and Northeastern University
Gina Green, Association of Professional Behavior Analysts

Stimulus generalization and contextual control affect the development of equivalence classes. Experiment 1 demonstrated primary stimulus generalization from the members of trained equivalence classes. Adults were taught to match six spoken Icelandic nouns and corresponding printed words and pictures to one another in computerized three-choice matching-to-sample tasks. Tests confirmed that six equivalence classes had formed. Without further training, plural forms of the stimuli were presented in tests for all matching performances. All participants demonstrated virtually errorless performances. In Experiment 2, classifications of the nouns used in Experiment 1 were brought under contextual control. Three nouns were feminine and three were masculine. The match-to-sample training taught participants to select a comparison of the same number as the sample (i.e., singular or plural) in the presence of contextual stimulus A regardless of noun gender. Concurrently, in the presence of contextual stimulus B, participants were taught to select a comparison of the same gender as the sample (i.e., feminine or masculine), regardless of number. Generalization was assessed using a card-sorting test. All participants eventually sorted the cards correctly into gender and number stimulus classes. When printed words used in training were replaced by their picture equivalents, participants demonstrated almost errorless performances.

Key words: stimulus equivalence, primary stimulus generalization, contextual control of class membership, generalization across tasks, verbal classes, language learning, humans

---

The study of language development has been dominated by the fields of linguistics and psycholinguistics. Research in these areas has provided valuable information about the structure of languages, about the verbal behavior that typically developing children display at different stages of language acquisition, and about the effects of a wide range of variables on language comprehension and production (Dale, 2004). In contrast, investigation of the necessary and sufficient conditions for producing language skills has been largely neglected (Dale, Roche, & Duran, 2008; Salzinger, 2008; Stemmer, 1990). For example, little is known about the conditions required to establish semantic and syntactic classes such as nouns, verbs, adjectives, singular, plural, and the like.

Recently Galizio, Stewart, and Pilgrim (2004) noted that ''It is important for beha-vior analysts to develop experimental paradigms for the production and analysis of complex, naturally occurring phenomena'' (pp. 254–255). A number of authors have suggested that Sidman's (1994) stimulus equivalence analysis is well-suited to the study of one such phenomenon, the development of verbal classes (e.g., Green, Sigurdar-dóttir, & Saunders, 1991; Griffee & Dougher, 2002; Houmanfar, Hayes, & Herbst, 2005; Lazar & Kotlarchyk, 1986; Mackay, 1991; Mackay & Fields, 2009; McIlvane, Dube, Green, & Serna, 1993; Sidman, 1986, 1994; Sigurdardóttir, Green, & Saunders, 1990). Sidman's (1971) original experiment demon-strated that training of certain conditional discriminations, involving pictures and their corresponding spoken and printed words with match-to-sample procedures, resulted in the emergence of other conditional discrimina-tions without direct training. As a result, the corresponding stimuli became mutually sub-stitutable, supporting the inference that each picture and its spoken and printed names constituted an equivalence class (Sidman,

1971, 1986, 1994; Sidman et al., 1982; Sidman & Tailby, 1982). This basic finding has now been replicated in scores of basic and applied studies involving various types of stimuli and participants, suggesting that stimulus equivalence plays an important role in the development of natural verbal classes.

Another likely contributor to the development of natural verbal classes is primary stimulus generalization, whereby stimuli become members of a class because they have physical characteristics in common with class members. They thus occasion the same response(s) as class members without further training. Several studies have examined interactions between primary stimulus generalization and equivalence class development and expansion (e.g., Adams, Fields, & Verhave, 1993; Barnes & Keenan, 1993; Fields, Reeve, Adams, Brown, & Verhave, 1997; Fields, Reeve, Adams, & Verhave, 1991; Galizio et al., 2004; Lane, Clow, Innis, & Critchfield, 1998). Results suggest that complex, verbal classes (or categories) may be accounted for by the combined effects of equivalence class formation and primary stimulus generalization. For example, the printed word DOG, the spoken word ''dog,'' and the picture of a golden retriever could become equivalent to, or substitutable for, one another as a result of conditional discrimination training and testing as in the Sidman equivalence model. Pictures of the same golden retriever taken from different angles and pictures of other physically similar dogs may also prove substitutable for the original pictures, thereby expanding the original equivalence class through primary stimulus generalization. Similar phenomena have been documented in several studies (e.g., Barnes & Keenan, 1993; Fields et al., 1991; Fields et al., 1997; Lane et al., 1998). To date, however, no published study has examined the development of entirely *new* stimulus classes based on primary stimulus generalization from members of existing equivalence classes.

In some natural languages, several words may be derived from a single word by the addition of a prefix or a suffix. For example, in English the addition of an ''s'' to some singular words makes the words plural (e.g., ''dog – dogs,'' ''cup – cups''). Equivalence and primary stimulus generalization procedures may provide means of experimentally analyzing the behavioral processes involved.

Suppose a non-English speaker is taught conditional discriminations leading to performances demonstrating that the spoken word ''dog,'' the printed word DOG, and the picture of a dog are equivalent. The individual might then demonstrate, without further training, that the spoken word ''dogs,'' the printed word DOGS, and a picture of several dogs are substitutable for one another. Indeed, those stimuli may also prove substitutable for the physically similar members of the original equivalence class, as a result suggesting that all six stimuli had formed one equivalence class.

For the English speaker in the example just described, stimuli present in typical conversational contexts probably determine whether any one of the six stimuli is treated as a member of one large equivalence class or as a member of the smaller classes that could be called (or tacted) ''singular'' and ''plural.'' Contextual stimuli probably function in several other ways to control membership in naturally occurring verbal classes. Consider an English speaker who has learned that the spoken word ''lightbulb,'' the corresponding printed word, and a picture of a lightbulb constitute an equivalence class. The individual then learns to match the picture of a lightbulb to the printed foreign-language word PERA, thus adding this word to the class. She is then likely to match PERA to the spoken and printed forms of the English word without direct training, and to say ''lightbulb'' in response to the question, ''What does 'pera' mean?'' Suppose further that this speaker also learns to match a picture of a pear to the spoken word ''pera.'' ''Pera'' thus may become a member of a second equivalence class containing the spoken word ''pear,'' the printed word PEAR, and the picture of a pear. Because the printed word PERA is now a member of two different equivalence classes, there is the potential for the two previously established equivalence classes to merge into one large class. When the speaker is asked ''What does 'pera' mean?'' she has no basis for responding (or may respond ''Either lightbulb or pear'') unless other cues are present to indicate which equivalence class is functional at the time (Bush, Sidman, & de Rose, 1989; Lynch & Green, 1991). When such contextual control is present, the occurrence of selective responding may be described in terms of class

intersection (Mackay, Wilkinson, Farrell, & Serna, 2011; Sidman, 1994).

In everyday discourse, several different types of contextual cues may determine the meaning—that is, the momentary equivalence class membership—of any given verbal stimulus. Some examples include physical surroundings, the topic of the conversation, and the vocal and nonvocal behavior of speakers and listeners (Sidman, 1986, 1994; Skinner, 1957). Several laboratory studies have demonstrated experimentally how contextual stimulus control can prevent the merger of separate equivalence classes into one, and determine whether a stimulus is treated as a member of one class or another. In those studies, contextual stimuli controlled the emergence of untrained conditional relations among stimuli that had never been presented with the contextual stimuli during training; that is, the contextual stimuli were shown to function independently as higher-order conditional stimuli, rather than as elements of compound sample (conditional) stimuli (Bush et al., 1989; Lynch & Green, 1991; Perez-Gonzalez & Serna, 2003; Serna & Pérez-Gonzalez, 2003; also see Sidman, 1986). A number of other studies also attempted to examine contextual stimulus control of equivalence class membership, but because the nominal contextual stimuli were paired with all of the stimuli in the experimenter-intended classes during both training and testing, the likelihood that they functioned as elements of compound sample and comparison stimuli rather than higher-order conditional stimuli cannot be ruled out (e.g., Gatch & Osborne, 1989; Griffee & Dougher, 2002; Hayes, Kohlenberg, & Hayes, 1991; Houmanfar et al, 2005; Kohlenberg, Hayes, & Hayes, 1991; Lazar & Kotlarchyk, 1986; Washio & Houmanfar, 2007). Stimuli designated as ''contextual'' in most studies have been nonlinguistic stimuli such as tones (Bush et al., 1989), spoken and printed nonsense syllables (Lynch & Green, 1991), abstract figures (Perez-Gonzales & Serna, 2003; Serna & Perez-Gonzalez, 2003), and colors (Griffee & Dougher, 2002; Hayes et al., 1991; Houmanfar et al., 2005; Lazar & Kotlarchyk, 1986; Washio & Houmanfar, 2007). However, Kohlenberg et al. (1991) used real linguisitic stimuli (proper English names generally considered to be ''male''

and ''female''). Mackay et al. (2011) used both English words for dog and bird pictures and also nonsense syllables and arbitrary forms in a set of experiments that experimentally assessed effects due to class merger and intersection. Further research is needed to illuminate the processes involved in treating naturally occurring linguistic stimuli as members of different stimulus classes depending on context, which seems central to aspects of language learning like categorization, semantics, and syntax (Mackay & Fields, 2009; Mackay et al., 2011).

The experiments described here used laboratory procedures to model the development of new verbal classes and contextual control of verbal class membership using real linguistic stimuli. Icelandic nouns were selected as experimental stimuli because they represent naturally occurring verbal classes that are well-suited to the study of primary stimulus generalization and contextual control. In Icelandic, nouns are either singular or plural, masculine or feminine. The first experiment used standard stimulus equivalence procedures combined with primary stimulus generalization to establish classes of Icelandic nouns in normally capable adults who had no prior knowledge of Icelandic or related languages. The second experiment aimed to establish contextual stimulus control of Icelandic noun class membership through higher-order conditional discrimination training and testing. Transfer of contextual control of class membership was then evaluated with a tabletop sorting task.

EXPERIMENT 1

Overview

This experiment served two purposes. First, we attempted to establish six 3-member auditory-visual equivalence classes, each consisting of a singular Icelandic spoken noun and its corresponding printed word and picture. After specific auditory-visual conditional relations were taught, tests were given to confirm the emergence of other conditional relations indicating that the trained relations had the properties that allow inference of equivalence (Sidman & Tailby, 1982). Oral reproduction/labelling tests then assessed whether participants produced the Icelandic

names of the singular pictures and printed words. Second, we examined the potential role of primary stimulus generalization in yielding new equivalence classes of stimuli that were plural forms of the original stimuli. The root of the singular and plural forms of each word were the same, and the plural versions of the pictures were multiples of the singular pictures, making it possible for classes of plural stimuli to emerge through primary stimulus generalization. For example, we asked if participants who demonstrated that the spoken singular noun ''bolli,'' the printed word BOLLI, and the picture of a cup were equivalent to each other would also match the spoken plural noun ''bollar,'' the printed word BOLLAR, and the picture of three cups to one another without explicit training.

METHOD

Participants

Four typical adults, (three females ages 26–54 years and one male age 30 years), and one typically developing 15-year-old girl participated. They were recruited through personal contact. Appendix 5 shows information about each participant's gender, age, native language, other languages spoken, etc. All were unfamiliar with Icelandic and related languages.

Apparatus, Setting, and Materials

A microcomputer and specially designed software (Dube, 1991) managed all experimental events (stimulus presentations, timing, response recording, etc.). Participants sat in front of the computer and responded to stimuli presented by the computer by using the mouse to position the cursor on the stimulus to be selected and clicking the mouse button. A Realistic amplifier-speaker system (model 32-2031A) connected to the computer presented all auditory stimuli. A tape recorder was used to record each participant's responses on oral reproduction/labelling tests.

Sessions were conducted in a quiet office with the computer positioned in the center of a desk. Participants sat in front of the computer with the mouse on a mouse pad to their right. They were instructed not to converse with the experimenter, who remained in the room but behind the participant.

Figure 1 shows the stimuli: dictated Icelandic nouns (set A) and their corresponding printed words (set B) and pictures (set C). Henceforth each stimulus in a set will be designated by the letter of that set and a number (i.e., from left in Figure 2 A1, A2, B3, B4, C1, etc.). The numbers denote the intended three-member equivalence classes (e.g., A1, B1, C1; A2, B2, C2).

Procedure

Figure 2 (upper panel) illustrates the initial training-testing paradigm. First, after tests for identity matching of pictures and printed words indicated that participants discriminated among the stimuli, six spoken word-printed word (AB) conditional relations were trained to criterion. For example, when ''bolli'' (A1) was the spoken sample, BOLLI (B1) was the correct printed word comparison. Then six spoken word-picture (AC) conditional relations were trained to criterion. For example, when ''bolli'' (A1) was the spoken sample, the picture of a cup (C1) was the correct comparison. Trial types are shown in Appendix 1.

Training. Each trial began with sample presentation. The sample key, a 3 cm × 5 cm area on the upper half of the computer screen, was black. The auditory sample was presented via the speaker and repeated every 3 s until the participant responded to the sample key. The first time participants encountered this task, they were instructed by the experimenter to ''Place the arrow on the black square and press the mouse.'' Following this observing response, presentations of the auditory sample ceased, the sample key turned white, and three comparison stimuli appeared in the lower half of the computer screen. The experimenter said, ''Touch one with the arrow and press the mouse.'' The verbal instructions were presented only in the first two trials of the first match-to-sample session.

Sessions were designed to teach designated conditional relations. For instance, in AB training, each trial had either A1, A2, or A3 as the sample, with B1, B2, and B3 as the comparisons. Samples changed unsystematically from trial to trial. The positions of the comparison stimuli changed from trial to trial,

SET						
A1-6	Spoken words	bolli	pera	fáni	kisa	brúsi	ugla
B1-6	Printed words	BOLLI	PERA	FÁNI	KISA	BRÚSI	UGLA
C1-6	Pictures	■	♡	⚑	☺	🖊	🦉

SET						
E1-6	Spoken words	bollar	perur	fánar	kisur	brúsar	uglur
F1-6	Printed words	BOLLAR	PERUR	FÁNAR	KISUR	BRÚSAR	UGLUR
G1-6	Pictures	■■■	♡♡♡	⚑⚑⚑	☺☺☺	🖊🖊🖊	🦉🦉🦉

Figure 1. Stimulus sets used in Experiment 1.

appearing equally often in each position, with the restriction that the comparison stimulus designated as correct did not appear in the same position on more than three consecutive trials. The AB and AC training each consisted of 36 trials. Performance criterion was 35/36 trials correct.

For training trials on which consequences were programmed (see below for details), responses to the comparison designated correct immediately produced a computer-generated jingle and a flashing screen. In addition, 1 cent was paid for each correct response and $2 for each completed daily

A

"bolli"   "fáni"   "brúsi"

"pera"   "kisa"   "ugla"

C

B

BOLLI   FÁNI   BRÚSI

PERA   KISA   UGLA

E

"bollar"   "fánar"   "brúsar"

"perur"   "kisur"   "uglur"

G

F

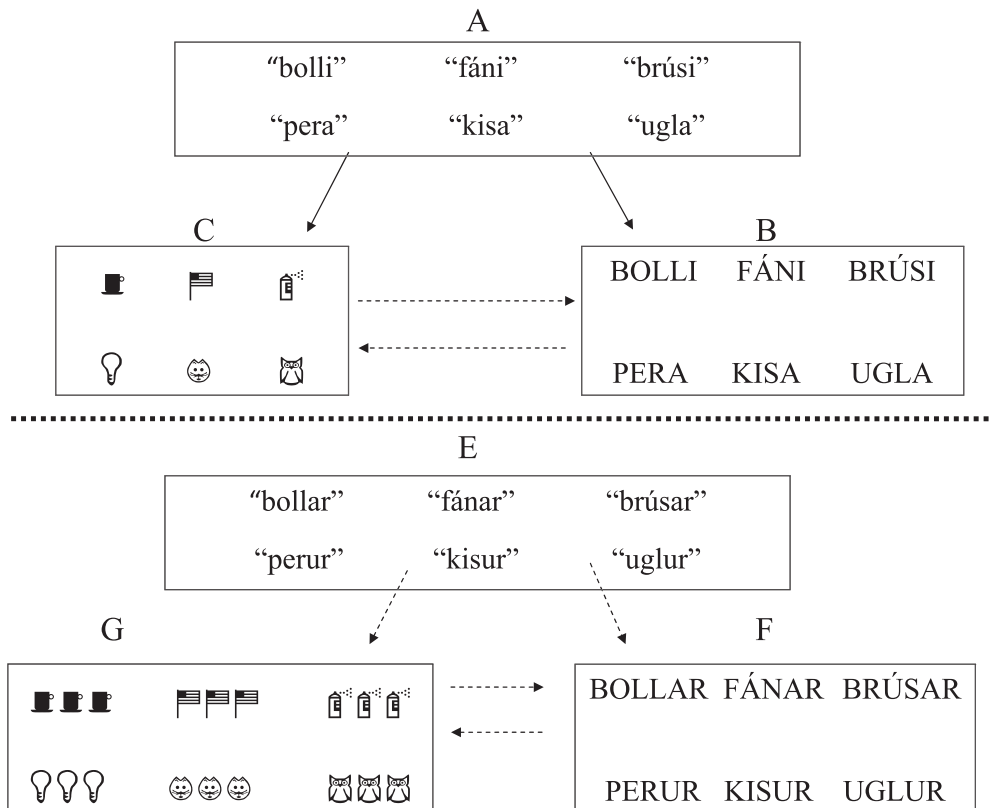BOLLAR FÁNAR BRÚSAR

PERUR KISUR UGLUR

Figure 2. Conditional relations trained (solid arrows) and tested (dashed arrows) in Experiment 1. Arrows point from sample to comparison stimuli.

session. Cash earned during the session was given to the participants after the session ended. A black screen of 1-s duration followed incorrect responses. The intertrial interval lasted 2 s during which time the screen was grey.

Training always started with programmed differential consequences following every trial. After criterion was met, reinforcement probability was reduced to .50. If performance was maintained at criterion, training trials were presented in extinction. During all apects of training, a record of correct responses was maintained for later use in calculating payment. If performance criterion was maintained in extinction, training began with the next set. If performance deteriorated below criterion during any session with reduced reinforcement, the probability of reinforcement was increased to the preceding level until performance criterion was demonstrated again at that level. Completion of training took a minimum of 216 trials. After performance met criterion in extinction, tests of untrained conditional relations began.

Tests for stimulus equivalence. Figure 2 (upper panel, dashed arrows) shows the conditional relations that were tested. For example, on the BC tests when the printed word BOLLI (B1) was the sample, the correct comparison was the picture of a cup (C1). On the CB tests, when the picture of a cup (C1) was the sample, the correct comparison was the printed word BOLLI (B1). These tests evaluated whether the relations established in training had the properties of symmetry and transitivity, critical properties of equivalence as described by Sidman & Tailby (1982). Trial types are shown in Appendix 1. A total of 72 test trials were given, 36 for each relation.

Each of these trials started with the presentation of the sample stimulus on the sample key, which was a visible gray square on the upper half of the screen. The first time participants encountered this task, they were instructed by the experimenter to ''Place the arrow on the square and press the mouse.'' This observing response was followed immediately by the appearance of three comparison stimuli in the lower half of the computer screen. The experimenter then said, ''Touch one with the arrow and press the

mouse.'' The sample remained on the sample key until a comparison stimulus was selected.

During test sessions, no programmed consequences followed any responses; each trial ended with removal of all stimuli (the screen showed only the sample key), and initiation of the intertrial interval. Criterion performance on these tests was the same as for training, that is, 35/36 trials. If criterion was not met, training and then tests were repeated.

Oral reproduction/labelling tests. Oral reproduction tests with picture stimuli and printed words were administered in that order. The purpose was to assess if the participants could produce orally the auditory stimuli used in the experiment or something so close to it that a speaker of Icelandic, naive to the experiment, would recognize the word the participant was trying to reproduce. The picture or printed word to be labelled appeared in the sample area on the computer screen. On the initial two or three trials of each test the experimenter said ''What is that called?'' following presentations of pictures, and ''What does it say?'' following presentations of printed words. Picture labelling tests were administered first, and word tests second. Each test consisted of 18 trials, with each stimulus presented three times in unsystematic order. All oral reproduction/labelling tests were audiotaped, and participants' responses were scored independently by two speakers of Icelandic. Participants proceeded to the generalization test regardless of performance on the oral reproduction/labelling tests.

Generalization test. Plural forms of the nouns used in training were presented on tests for generalization. The stimulus sets used in these tests are shown in Figure 2 (lower panel): spoken words (set E); printed words (set F); and pictures (set G). Figure 2 (lower panel, dashed arrows) shows untrained relations that were tested in this condition. For example, on the FG test, when the printed word BOLLAR (F1) was the sample, the picture of three cups G1) was the correct comparison, and on the GF test, when the picture of three cups (G1) was the sample, the printed word BOLLAR (F1) was the corresponding correct comparison. Trial types are shown in Appendix 1. Each test consisted of 36 trials.

Oral reproduction/labelling tests of plural pictures and words. Participants were given oral reproduction/labelling tests with the

plural forms of the pictures and printed words, in that order. These tests were done in the same way as the previous oral reproduction tests described above and had the same purpose.

Inter-scorer agreement for performances in oral reproduction/labelling tests. Audiotaped performances of participants M.T., T.M., and R.C. were used to assess interobserver agreement on oral reproduction/labelling performances. First, after the principal experimenter (who spoke Icelandic) rated these 3 participants performances on all oral reproduction/labelling tests, another native speaker of Icelandic who was naive to the experimental hypotheses and conditions did the same. He merely listened to the tapes and wrote down the words he heard the participants say. A second rater, also naive to the experimental hypotheses and conditions, listened to the tapes of all participants but used trial-by-trial lists of the words and pictures presented. This observer recorded whether the participant said the same word as that listed for the particular trial, or, when that did not occur, wrote the word that he heard the participant say. Interobserver agreement was calculated between the experimenter and the first rater, then separately between the experimenter and the second rater. In both cases interobserver agreement was calculated by totalling the number of times that the scorers

agreed that the participant said or did not say a particular word (trial by trial), dividing that total by the number of words in each oral reproduction/labelling test (18), and multiplying by 100%. Interobserver agreement scores can be seen in Table 1. Interobserver agreement between experimenter and Scorer 2 was almost always higher.

RESULTS

Training

Table 2 summarizes the results. All participants demonstrated criterion performance on all AB relations in a single 36-trial training set. The performances of 3 participants were errorless. Two participants (M.B. and R.C.) made one error. All performances remained at criterion during reduction of reinforcement and during the extinction phase of training. All participants required two exposures to the AC training with programmed consequences on every trial to achieve criterion performance. Performances remained at criterion during reduction of reinforcement and during extinction.

Equivalence Tests

On CB tests, where pictures were samples and printed words were comparisons, all

Table 1
Percentage of Interobserver Agreement in Oral Reproduction/Labelling Tests

Participant	Pictures	Words	Pictures	Words	Mean
Scorer 1					
M.T.	94	67	89	83	83
T.M.	94	83	89	83	87
R.C.	100	78	89	61	82
					$\bar{X} = 84$
Scorer 2					
M.T.	94	83	83	83	86
T.M.	100	83	89	100	93
R.C.	100	100	89	72	90
					$\bar{X} = 90$
C.S.	83	89	94	94	90
M.B.	89	94	83	72	84.5
					$\bar{X} = 89$

Note. Experimenter's ratings were compared to the ratings of Scorer 1 who did not know what words the participants were trying to reproduce and then to the ratings of Scorer 2 who knew.

Table 2

Participants' Performances in AB and AC Training, on Tests for Equivalence Class Formation, and on Oral Reproduction/Labelling Tests in Experiment 1

| Participant | Training | | | | | | Tests | | Oral reproduction/ labelling | |
| | AB % reinforcement | | | AC % reinforcement | | | | | | |
	100	50	0	100	50	0	CB	BC	Pictures	Words
M.B.	35	36	36	34/36	36	35	36	34/36	12	12
R.C.	35	36	36	30/36	36	36	33/35	36	8	9
T.M.	36	36	36	26/35	36	36	36	36	13	18
M.T.	36	36	36	28/35	36	36	36	36	10	18
C.S.	36	36	36	25/35	36	36	36	36	8	9

Note. Numbers in table refer to the number of correct trials (of a total of 36 match-to-sample trials and 18 oral reproduction/labelling trials).

#/# indicates scores on 1st and 2nd test administrations respectively.

participants performed errorlessly except participant R.C. (see Table 2, second row, CB column). She made three errors, consistently selecting the word ''brusi'' (spray can) in response to the picture of the owl. On the BC test where the printed words were samples and the pictures were comparisons, however, R.C. made no errors (see second row, BC column). When she was given the CB test again, R.C. did not repeat the errors she made earlier. She made one error on a different trial type than before, but her performance met criterion.

During the BC test, where printed words were samples and pictures were comparisons, four participants had errorless performances. Participant M.B. made two errors during the first administration of this test, none in the second.

Oral Reproduction/Labelling Tests

Results of oral reproduction/labelling tests are shown in the rightmost columns of Table 2. Most participants were consistent in labelling or producing the names of certain pictures and words and not others. Detailed analyses of the naming performances are available from the first author upon request. Participants labelled 8–13 of the 18 pictures with corresponding Icelandic names. Subject R.C. said on some trials that she could not replicate the sound she heard from the computer, but stated that she knew ''how the name for that picture is spelled'' and spelled the Icelandic name aloud correctly.

Participants reproduced 9–18 of the 18 printed words. On these trials R.C. claimed to be unable to replicate the Icelandic words she had heard during training but offered the English equivalents. For example, upon seeing the word BOLLI she said, ''That means cup.'' Then she added that if she read the letters the way that was natural to her it would ''not sound the same way that the computer says it.''

Generalization Tests

Table 3 summarizes the results of the four generalization tests. All performances met criterion. Many performances were perfect, and no participant made more than one error on any trial type.

Oral Reproduction/Labelling of Plural Pictures and Printed Words

Results of the oral reproduction/labelling tests are shown in Table 3. Mean interobserver agreement was 84% (with Scorer 1)–90% (with Scorer 2). Scores ranged from 5–11 correct. Participants M.B., T.M., and M.T. gave the singular Icelandic word for some of the plural pictures. For example, upon seeing the picture of three cats they said ''kisa'' rather that ''kisur,'' and upon seeing the picture of three owls they said ''ugla'' rather than ''uglur.'' Participant R.C., whose first language was English, used the English plural ending of some words, that is, reliably adding an ''s'' to the singular Icelandic

Table 3

Participants' Performances on Tests for Stimulus Generalization and Oral Reproduction/ Labelling in Experiment 1

Participant	Generalization tests				Oral reproduction/labelling	
	GF	FG	EG	EF	Pictures	Words
M.B.	34^a	36	36	36	$12(3^c)$	$17(2^c)$
R.C.	34^a	36	35	36	$5(4^b)$	13
T.M.	35	35	36	36	$11(6^c)$	18
M.T.	35	36	36	36	$7(3^c)$	$10(1^c)$
C.S.	35	36	34^a	35	5	9

Notes. a = one error in each of two relations; b = errors, gave English plural form (-s); c = errors, used singular form.

Numbers in table refer to the number of correct trials (of a total of 36 match-to-sample trials and 18 oral reproduction/labelling trials).

name. For example, upon seeing the picture of three cats she said ''kisas'' rather than ''kisur.'' She also said ''perus'' rather than ''perur'' in response to the picture of three lightbulbs.

The participants' accuracy in reproducing plural printed words ranged from 9–18 correct. Participants M.B. and T.M. reproduced all plural printed words correctly. Participant M.T. gave the singular name ''ugla'' in response to the printed plural word UGLUR in the same way as he had done earlier upon seeing the picture of three owls.

DISCUSSION

The results of Experiment 1 add to the literature demonstrating that conditional discrimination training can generate new conditional relations that were not trained directly. Participants performed better on matching printed Icelandic words to corresponding spoken words (AB training) than matching pictures to those same spoken words (AC training). Most participants needed only one exposure to AB training to achieve criterion performance, whereas they needed two exposures to achieve criterion performance on the AC relations. Participants' knowledge of phonics may have aided their acquisition of the AB relations. Of course, that history would not have enabled them to match picture comparisons to spoken Icelandic word samples.

The results of the oral reproduction/labelling tests suggest that the emergence of untrained conditional relations may be independent of oral reproduction of auditory sample stimuli. The oral reproduction/labelling performances of all participants in this experiment were poor (according to native speakers of Icelandic), ranging from 45%– 70% on the picture naming test. Performances on printed word labelling tests were generally better, but not perfect. Also, 1 participant did not label the pictures but spelled their names instead. In this experiment then, reproducing auditory labels of pictures and printed words was not a necessary outcome of formation of equivalence classes that included auditory stimuli. The performance of participant R.C. on oral reproduction/labelling of plural pictures and words differed from the performances of the other participants. She used plural English word endings to form the plurals of the Icelandic words. This was likely attributable to her extensive prior experience with English, her first language. Her behavior was similar to that shown by young children learning to form English plurals, who often apply ''s'' to singular nouns even when that is inconsistent with the rules of English. For example, Berko (1958) showed that children exposed to English morphology constructed plurals by adding ''s'' to nonsense words like ''wug.'' This class of behavior is typically referred to as ''overextension'' in nonbehavior analytic research on language development (e.g., Gruendel, 1977; Thomson & Chapman, 1977).

Results of the generalization tests suggest that primary stimulus generalization yielded

performances in addition to those that are typically demonstrated on tests for the properties of equivalence. Participants matched plural spoken words to corresponding printed words and pictures without direct training. Primary stimulus generalization apparently enabled participants to respond appropriately in this novel situation.

It was not clear whether the generalization test performances were evidence of the expansion of classes established during training, or of the formation of new classes that included only plural stimuli. Tests for dicrimination of plural and singular stimuli that would clarify this point were not administered because participants often labelled the stimuli in the generalization tests ''plural'' as soon as they were exposed to them. In order to avoid fostering a bias toward assigning all stimuli to either a class of singular or a class of plural stimuli, the tests for dicrimination of singular from plural stimuli were postponed. Training in Experiment 2 was aimed at establishing contextual control of classes, one aspect being number (i.e., singular vs. plural context). Thus, a preexisting bias that might induce control by mere exposure to test trials, had to be avoided.

EXPERIMENT 2

Overview

Experiment 2 had five aims. The first was to establish contextual stimulus control over certain trained conditional relations among the stimuli used in Experiment 1, as illustrated in Figure 3. As noted earlier, Icelandic nouns are classified as singular or plural, masculine or feminine. In the presence of the printed word TALA (''number'' in Icelandic) as a contextual stimulus, printed singular noun samples were to be matched to other printed singular noun comparisons, and plural printed noun samples were to be matched to other plural printed noun comparisons; the gender of the words was irrelevant on those trials. On other trials within the same training sessions, the contextual stimulus was the printed word KYN (''gender'' in Icelandic). In that context, masculine printed noun samples were to be matched to other masculine printed noun comparisons, and feminine printed noun samples were to be matched to

other feminine printed noun comparisons; number was irrelevant.

The second aim was to assess the effects of contextual control training on the demonstration of classes of singular, plural, feminine, and masculine nouns in a card-sorting task. The stimulus groupings expected of the computerized contextual control training transferred to the sorting task are shown in the upper panel of Figure 4.

Third, we tested for discrimination between singular and plural stimuli in a match-to-sample context. (This test could have been given at the end of Experiment 1 but was not given in order to avoid creating a difficult-to-modify bias before starting Experiment 2.) In these test trials participants were presented with a singular or a plural picture (e.g., a cup) and had to select a printed singular or plural word (i.e., cup vs. cups). We asked whether they would respond to corresponding plural printed names in the presence of plural picture samples and to singular corresponding printed names in the presence of singular picture stimuli. If so, it would be possible to conlude that classes of plural stimuli had emerged without specific training due to primary stimulus generalization.

The last two purposes of this experiment were first to assess the effects of substituting equivalent stimuli in contextually controlled conditional relations. These trials were arranged like contextual control training trials, except that the printed nouns used as samples and comparisons in that training were replaced with corresponding pictures from the equivalence classes developed in Experiment 1. Finally, generalization to the sorting test was examined. The task involved sorting 12 cards on which the pictures from this phase were presented, along with 12 cards on which the printed nouns were presented, in the presence of each of the contextual stimuli (Figure 4, lower panel).

Stimuli

Table 4 shows the stimuli used in this experiment. The printed words from Experiment 1 were reorganized into different sets as follows: set O—singular masculine nouns; set P—singular feminine nouns; set Q—plural feminine nouns; and set R—plural masculine nouns. Pictures that were in the
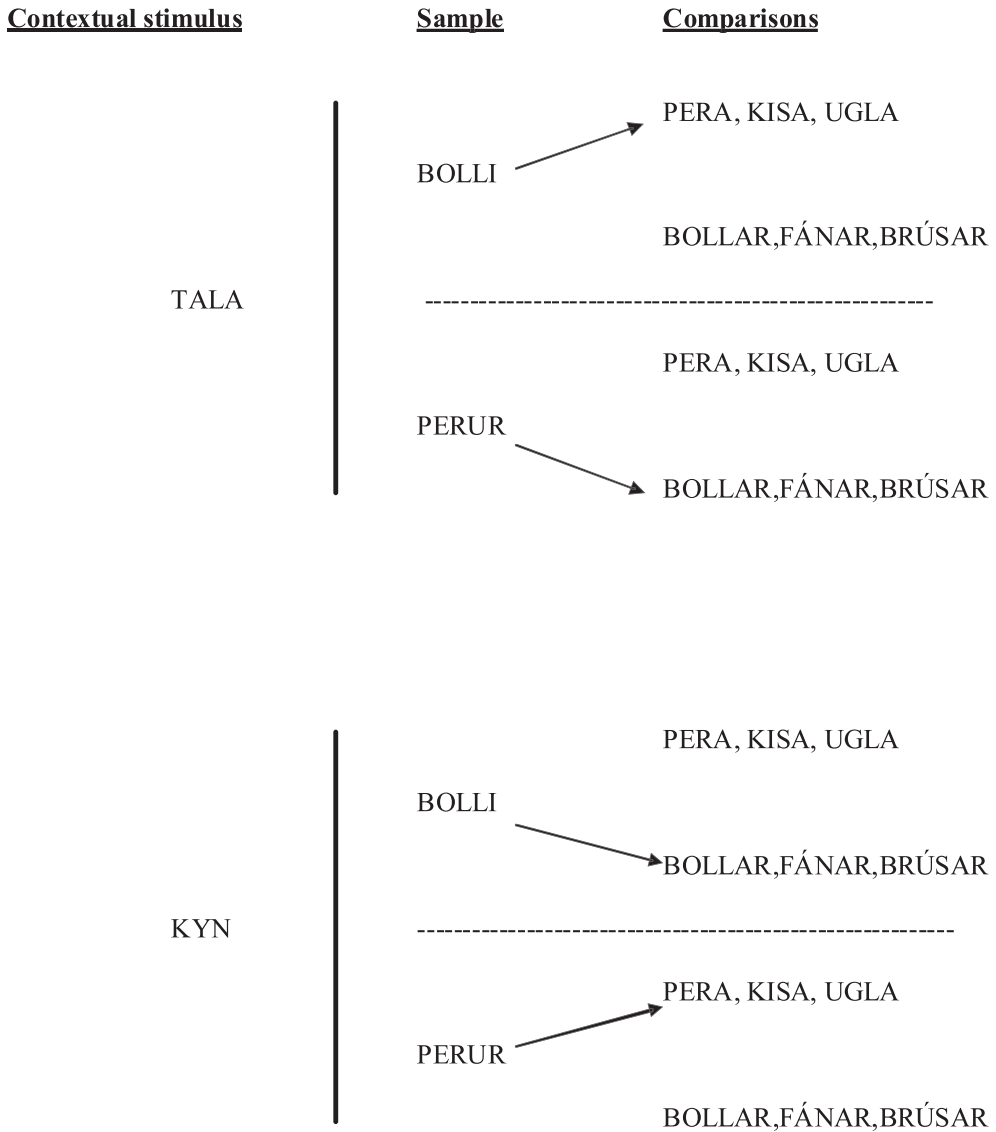
Contextual stimulus   **Sample**   **Comparisons**

PERA, KISA, UGLA

BOLLI

BOLLAR,FÁNAR,BRÚSAR

TALA   --

PERA, KISA, UGLA

PERUR

BOLLAR,FÁNAR,BRÚSAR

PERA, KISA, UGLA

BOLLI

BOLLAR,FÁNAR,BRÚSAR

KYN   ---

PERA, KISA, UGLA

PERUR

BOLLAR,FÁNAR,BRÚSAR

Figure 3. Representative contextually controlled conditional relations established in Experiment 2. Arrows point from samples to comparisons that were designated correct with those samples in the presence of each of the contextual stimuli, the printed words TALA (''number'' in Icelandic) and KYN (''gender'' in Icelandic).

Experiment 1 equivalence classes with the printed stimuli were reorganized into sets K, L, M, and N.

Stimuli were presented via computer (as described previously, with the addition of the higher-order or contextual stimuli) for contextual control training. For sorting tasks, replicas of the computer-presented stimuli were laser-printed on white 13 cm × 8 cm cards.

Procedure

Immediately after finishing Experiment 1, participants began Experiment 2 with the training illustrated in Figure 3. In the presence of TALA and the printed noun BOLLI as sample, responses to comparison stimuli PERA, KISA, and UGLA were correct; all of those nouns are singular but feminine. On

	TALA		KYN	
SINGULAR	PLURAL	MASCULINE		FEMININE
BOLLI	BOLLAR	BOLLI		PERA
FÁNI	FÁNAR	BOLLAR		PERUR
BRÚSI	BRÚSAR	FÁNI		KISA
PERA	PERUR	FÁNAR		KISUR
KISA	KISUR	BRÚSI		UGLA
UGLA	UGLUR	BRÚSAR		UGLUR

	TALA		KYN	
SINGULAR	PLURAL	MASCULINE		FEMININE
BOLLI ▪	BOLLAR ▪▪▪	BOLLI ▪		PERA ♀
FÁNI ▤	FÁNAR ▤▤▤	BOLLAR ▪▪▪		PERUR ♀♀♀
BRÚSI ▤	BRÚSAR ▤▤▤	FÁNI ▤		KISA ☺
PERA ♀	PERUR ♀♀♀	FÁNAR ▤▤▤		KISUR ☺☺☺
KISA ☺	KISUR ☺☺☺	BRÚSI ▤		UGLA ☒
UGLA ☒	UGLUR ☒☒☒	BRÚSAR ▤▤▤		UGLUR ☒☒☒

Figure 4. Stimuli for Experiment 2 sorting tasks. Each stimulus card showed one printed word or a picture of either one or three items.

other trials with TALA as the contextual stimulus and a plural printed noun (e.g., PERUR) as sample, responses to comparison stimuli BOLLAR, FÁNAR, and BRÚSAR were correct; all of those nouns are plural but masculine. In contrast, in the presence of KYN as the contextual stimulus and a printed noun (e.g., PERUR) as the sample, responses to comparison stimuli PERA, KISA, and UGLA were corect because all are feminine singular nouns. Finally, in the presence of KYN and a printed noun (e.g., BOLLI), selecting BOLLAR, FÁNAR, and BRÚSAR was correct because these are all masculine plural nouns.

Contextual control training. Contextual control of conditional relations among printed words was trained with computerized, two-choice, match-to-sample procedures. Each trial started with presentation of a printed word contextual stimulus in the top part of the sample area. An observing response to the contextual stimulus produced a sample directly under the contextual stimulus. Samples were only from stimulus sets O an Q (see in Table 4) in order to allow testing for emergent relations with stimuli from sets P and R. An observing response to the sample was followed immediately by the appearance of two comparison stimuli at the bottom of the screen, in two out of three possible positions. These positions varied unsystematically from trial to trial as described earlier. Programmed consequences also were as described before. The 12 trial types are shown in Appendix 2. To train each contextually controlled conditional discrimination took 108 trials. Two were trained simultaneously. Three sets (each with 216 trials) were needed to train all contextually controlled conditional relations with all stimuli from sets O and Q as samples. After criterion performance was

Table 4
Stimuli in Experiment 2

TALA	Singular masculine	Set O: BOLLI, FÁNI, BRÚSI	Set K:
KYN	Singular feminine	Set P: PERA, KISA, UGLA	Set L:
	Plural feminine	Set Q: PERUR, KISUR, UGLUR	Set M:
	Plural masculine	Set R: BOLLAR, FÁNAR, BRÚSAR	Set N:

achieved (no more than one error on any contextually controlled conditional relation, i.e., in 108 trials), the participant progressed to the next experimental condition.

Sorting test—Printed nouns. This test evaluated whether participants would sort the printed nouns into classes of feminine, masculine, singular, and plural depending on the contextual stimulus presented. In other words, it assessed transfer of the computerized contextual control training to a noncomputerized, tabletop sorting task. Figure 4 (upper panel) illustrates the sorting expected if transfer occurred. In the presence of contextual stimulus TALA (number), all singular words would be sorted into one set and all plural words into another set (upper panels). Additionally, in the presence of contextual stimulus KYN (gender), all masculine words would be sorted into one set and all feminine words into another. Pictures, which were presented in another test described later, would also be sorted into singular and plural sets in the presence of TALA and into masculine and feminine sets in the presence of KYN (lower panels).

The sorting test also assessed equivalence relations among the stimuli in the different categories. Sorting all singular words into a single set in the presence of TALA is consistent with the notion that the trained relations between those stimuli (shown in Figure 4) were symmetric and transitive. To illustrate, symmetry of the trained conditional relation between BOLLI (sample) and PERA (comparison) would involve selecting BOLLI as the correct comparison given PERA as a sample. Similar logic applied to plural words in the presence of TALA and masculine and feminine words in the presence of KYN. Addtionally, because the printed words and their corresponding pictures had been demonstrated to constitute equivalence classes, the sorting outcomes illustrated in the lower panel of Figure 4 were predicted. Such sorting tests have been used to assess equivalence class formation in previous research reported, for example, by Cowley, Green, and Braunling-McMorrow (1992), Lowe, Horne, Harris, and Randle (2002), Mackay et al. (2011), and Pilgrim and Galizio (1996).

To start each sorting trial, the participant was given 12 cards, each presenting one of the printed nouns. The experimenter placed a contextual stimulus printed on a similar card on the table and said, ''Arrange the cards in your hand in any way you want depending on which cards you think go together when you have this card (pointing at the contextual stimulus) in front of you.'' Each contextual stimulus was presented three times in unsystematic order for a total of six sorting trials. Criterion performance was at least one perfect sorting into the groupings shown in Figure 4 with each contextual stimulus. No programmed consequences were provided at any time. If criterion was not met within the six opportunities provided, remedial training began. After remedial training, sorting tests were given again. Participant-specific procedural details of remedial training and testing will be reported with the results.

Singular vs. plural discrimination test. This computerized match-to-sample test evaluated discriminaton of singular and plural printed words and pictures on 72 trials with no programmed consequences. All trial types are listed in Appendix 3. Each trial presented a singular picture sample and three printed noun comparisons, one singular noun and either another singular noun and a plural noun or two plural nouns (comparison types were distributed equally across trials). For example, on a trial when the picture of a cup was the sample, comparisons were the printed words BOLLI, BOLLAR, and either FÁNI or FÁNAR. On other trials, plural picture samples were presented with the same comparison arrays just described. If all the singular and plural nouns were in the same classes, participants would be as likely to respond to singular as plural printed nouns in the presence of either singular or plural picture samples. If they responded to plural printed nouns in the presence of plural picture samples and to singular printed nouns in the presence of singular pictures, it would indicate that classes of plural stimuli had emerged without specific training due to primary stimulus generalization. Alternatively, mere exposure to these test trials could suffice to establish classes of plural stimuli. These trials were the first occasions on which the singular and plural versions of the same word were presented simultaneously as comparisons. Because the procedure permitted selection of only one of the stimuli, these trials may have fostered discriminating between singulars and plurals.

Contextual control test with pictures. In this condition, we presented the same type and number of trials as in the contextual control training, except that printed nouns were replaced by pictures from the Experiment 1 equivalence classes. Trial types are shown in Appendix 4.

Sorting test—Printed nouns and pictures. On this test, the participant was given 24 cards: the 12 printed nouns and the 12 pictures shown in the lower panel of Figure 4. Otherwise, procedures and the performance criterion were as described for the first sorting task.

RESULTS

Contextual Control Training

Table 5 shows that all participants learned the contextually controlled conditional relations.

Table 5

Performances of Participants in Training of Contextually Controlled Conditional Discriminations

Participant	Set 1	Set 2	Set 3
M.B.	169, 185, 194, 194, 198, 210 (1296)	205, 212 (432)	200, 213 (432)
R.C.	151, 162, 148, 170, 183, 174, 176, 213 (1728)	175, 172, 172, 169, 216 (1080)	213 (432)
T.M.	208, 106/108* (324)	213 (216)	215 (216)
M.T.	181, 212 (432)	213 (216)	210 (216)
C.S.	150, 131, 172, 209, 107/108* (972)	187, 215 (432)	213 (216)

Notes. There were 216 trials in each training set. Number of correct trials are shown. Numbers in parenthesis show trials to criterion.

* = performance on half of the set, criterion met.

The list of participants is in an order consistent with Experiment 1 where acquisition data differed little among participants. Here, substantial differences were observed in the amount of training required, particularly for Set 1. Participants T.M. and M.T. learned the first group of contextually controlled conditional relations in only one and two exposures to Set 1, respectively. Aquisition of the second and third groups of contextually controlled conditional relations required only one exposure to each training set. Participant C.S. repeated the first training set four times before she started responding consistently and correctly. She subsequently needed two exposures to the second training set and one exposure to the third training set for her performance to meet criterion. Participants M.B. and R.C. required even more exposures (6 and 8 respectively) to the first training set to attain criterion performance. Participant M.B. then needed two exposures to each of the second and third training sets for performance to meet cirterion. Participant R.C. needed five exposures to the second training set and one exposure to the third training set to attain that level.

Sorting Test—Printed Nouns

Participant R.C. was the only one who sorted all the singular printed nouns together and all the plural nouns together, regardless of gender, in the presence of the contextual stimulus TALA (number) and all the masculine nouns and all the feminine nouns into separate sets, regardless of number, in the presence of KYN (gender). She did so on two consecutive trials. Participant M. B. sorted all feminine and masculine words together in the presence of KYN, but failed three times to sort correctly in the presence of TALA. Participants M.T. and T.M. sorted all singular and plural words together in the presence of TALA, but failed three times to sort the cards according to gender in the presence of KYN. Participant C.S. did not arrange the stimuli in any systematic way.

Singular vs. Plural Discrimination Test

As shown in Table 6 (third major column from left), all participants matched singular printed nouns to singular picture samples and

Table 6

Performances on Tests Following Contextual Control Training. Numbers Refer to the Number of Trials Given Before Criterion Was Met.

Participant	Sorting test		Test singular vs. plural classes	Sorting test repeat		Test contextual control with pictures	Sorting test (words & pictures)
	TALA	KYN		TALA	KYN		
M. B.	FFF	PPP	P	P	P	756	P
R. C.	P	P	P	N/A	N/A	648	P
T. M.	PPP	FFF	P	P	FF	648	F
M. T.	PPP	FFF	P	P	FFF	648	F
C. S.	FFF	FFF	P	N/A	N/A	648	PPP, FFF

P = passed, F = failed.

plural printed nouns to plural picture samples. These results indicate that classes of plural stimuli (printed nouns and corresponding picture sets) emerged from the equivalence class training and testing in Experiment 1. The likely basis for the development of the plural classes was primary stimulus generalization, in that the plural forms of the printed nouns and pictures had physical featues in common with the singular printed nouns and pictures that constituted equivalence classes in Experiment 1.

Sorting Test—Repeat

Table 6 shows performances on repetiton of the sorting test, following the singular vs. plural discrimination test. Participant M.B., who had failed to sort the cards with respect to number in the presence of contextual stimulus TALA, now did so. Participants T.M. and M.T., who had failed to sort the cards with respect to gender in the presence of KYN, failed to do so again but continued, as before, to sort correctly in the presence of TALA.

Contextual Control Tests With Pictures

Results of these tests are shown in Table 6. All participants responded to the picture comparisons in the presence of picture samples and the contextual stimuli just as they had been trained to respond when the equivalent printed nouns were samples and comparisons. Very few errors occurred.

Sorting Test—Printed Nouns and Pictures

Table 6 (right column) shows that participants M.B. and R.C. sorted the 12 picture cards and the 12 printed word cards correctly in the presence of the contextual stimuli, just as they had previously. Participants M.T., T.M., and C.S. sorted the pictures and cards correctly in the presence of contextual stimulus TALA but not in the presence of KYN. Participant C.S. sorted the singular picture and word cards into one pile and the plural picture and word cards into another pile in the presence of both contextual stimuli.

Remedial Training and Testing

In an attempt to remediate the failure by participants M.T., T.M., and C.S. to sort the

cards correctly in the presence of contextual stimulus KYN (gender), they were all given two reviews of contextual control training with printed words, each followed by a sorting test with words only. Contextual control training with pictures rather than printed words was added if the previous reviews failed to yield correct sorting. Performances for each individual are shown in Table 7.

T.M. (top line) passed the sorting test with printed word stimuli after the second review of contextual control training. She then was given the sorting test with both printed word and picture stimuli, but failed when the contextual stimulus was KYN. Next, she was given the sorting test with printed word stimuli only, a test she had passed previously. On this occasion, she passed the sorting test with printed words again, and then did so with both printed words and pictures. Repeated testing thus sufficed to yield performance generality.

M.T. (middle row in Table 7) failed the sorting tests given after the first and second reviews of contextual control training. He then was given contextual control training but with picture stimuli instead of the printed words used earlier. He then passed the sorting test with both words and pictures.

C.S. (bottom row in Table 7) received the same remedial training as M.T. but still failed the sorting test. Therefore, she was given a test in which she was exposed for the first time to match-to-sample trials like the following: Given the contextual stimulus KYN and a singular-masculine sample (e.g., BOLLI), the comparison array included two singular printed word comparisons, one of which was feminine and the other masculine, thus emphasizing the gender discrimination. Other trials presented the contextual stimulus KYN with a plural feminine sample and two plural comparisons, one feminine and one masculine, again emphasizing gender. Initially C.S.'s responses on these trials appeared random, but she began responding as predicted after 324 trials and then continued to do so. She passed the sorting tests after this remedial training and verbalized for the first time during the experiment the rules by which she sorted the cards, stating that ''Brúsi goes with brúsar but brúsi also goes with fáni, so it can also go with fánar and the same is for bolli and bollar, so all these [putting all masculine words

Table 7

Performances of Participants T.M., M.T., and C.S. in Remedial Training and Tests in Experiment 2. Numbers Refer to Total Number of Trials in the Training or Tests Provided.

Participant	Sorting test—words		Sorting test—words		Sorting test—words & pictures		Sorting test—words		Sorting test—words & pictures	
	TALA	KYN	TALA	KYN	TALA	KYN	TALA	KYN	TALA	KYN
T.M.	P	F	P	P		F	P	P	P	P
						648				

Participant	Sorting test—words		Sorting test—words		Train contextual control, pictures	Sorting test—words		Sorting test—words & pictures	
	TALA	KYN	TALA	KYN		TALA	KYN	TALA	KYN
M.T.	P	F	PPP	FFF	648	P	P	P	P

Participant	Sorting test—words		Sorting test—words		Test symmetry & transitivity under contextual control		Sorting test—words		Sorting words & pictures	
	TALA	KYN	TALA	KYN	Equiv.	Trans.	TALA	KYN	TALA	KYN
C.S.	P	F	P	F	864	864	P	P	P	P
				FFF						
				648						

together] can go together. And it is the same for all these [placing all the feminine word cards together].''

DISCUSSION

All participants acquired the contextually controlled conditional relations after varying amounts of higher-order match-to-sample training. Acquisition generally proceeded more quickly with stimulus sets presented later in training. After the computerized training had established contextual control of participants' match-to-sample performances, only R.C. sorted the printed nouns and pictures under contexual control in a way that was consistent with training. It appeared that only one of the two contextual stimuli controlled the sorting performances of the other participants, but the particular stimulus that did so differed across participants. In short, the contextual stimuli clearly controlled differential behavior in sorting tasks, but not necessarily as intended by the experimenter. Participant M.B. sorted correctly with respect to gender but not with regard to number until after he was exposed to the test requiring discrimination of singular and plural nouns. In contrast, participants T.M. and M.T. sorted correctly with respect to number but not gender until contextual control training was repeated several times. Participant C.S. did not sort the stimuli in any systematic way until contextual control training as well as tests for contextually controlled emergent relations were repeated.

The effects of repetition and change of procedure across training and testing sessions are not well understood. However, these factors may affect performance in extended experiments such as the current ones, indicating to some participants, for example, that their performance was not yet ''correct.'' Historically, each time they produced a consistent response pattern during training and testing, the experimenter introduced either new stimuli (e.g., the next phase of training following criterion performance), or new contingencies (e.g., extinction on tests following training). In contrast, the occurence of prolonged or repeated testing may have implied that the ''correct'' pattern had not been produced. Under these conditions, the result may be continued production of new response patterns until testing stops (cf. Saunders & Green, 1992). The continuation

or change in procedures thus may serve as an ''instruction'' relevant to performance on a subsequent set of trials (cf. Sidman, Wynne, Maguire, & Barnes [1989]). It is possible that the effects observed under the conditions described here would be less pronounced in nonexperimental situations where the relevant changes in contingencies are likely to occur more irregularly.

The results obtained with participants T.M. and M.T. add to the body of evidence showing that equivalence classes can be very resistant to change. Repeated contextual control training and presentation of the sorting tasks were required before these participants demonstrated the behavior intended by the experimenter (cf. Green et al., 1991; Saunders, Saunders, Kirby, & Spradlin, 1988; Saunders, Wachter, & Spradlin, 1988; Sidman et al., 1989). Participant M.T. spontaneously produced oral names for two stimuli during conditional discrimination training. Those names may have contributed to the development of equivalence classes or stimulus control topographies (Dube & McIlvane, 1996) other than those intended by the experimenter, and which subsequently proved relatively insensitive to experimental contingencies (Catania, 1992).

It is also worth noting that participants T.M. and M.T. passed the sorting task only after they received extensive computerized contextual control training and/or testing comparable to that required originally R.C. and M.B. (see Table 5). This raises the possibility that amount of exposure may affect transfer to related tasks. For M.T. and T.M., who acquired the contextual control performances relatively quickly, additional exposure (a form of overtraining) appeared to have been necessary to produce correct performances on the sorting test. In the cases of R.C. and M.B., such additional training was not necessary. Their earlier extensive exposure to contextual control requirements sufficed to permit transfer to the new task format, this raises questions regarding training criteria used in conditional discrimination and stimulus equivalence research. Often 90–95% accuracy on all conditional relations trained for one or two sessions is required for participants to proceed to the next phase in equivalence experiments. That may not be sufficient to produce predicted outcomes on tests for equivalence or tests for

generalization; maintenance of criterion performance for some extended period may be a critical, even necessary, prerequisite. On the other hand, the provision of extended training was not sufficient to produce contextually controlled sorting performance by C.S., the youngest participant.

A variable that might account for some of the differences in performances across participants is the presence of gender differentiation in their native languages. Such a history could enable discrimination of analogous dichotomous structures in other languages. Interestingly, the 2 participants who performed correctly in the card-sorting test immediately after contextual control training (R.C. and M.B.) were native speakers of languages in which gender differentiation is not present. R.C. spoke only English and M.B.'s native language was Armenian. M.B. was a beginning student of English. Moreover, the 2 participants who received extensive remedial training (C.S. and M.T) were both native speakers of languages in which gender differentiation is present (Portuguese and Arabic, respectively). The third participant who needed remedial training before demonstrating correct performance on the card-sorting test was a native speaker of Malayalam, a language spoken in a southern state of India, in which gender differentiation is not present. She had spoken English for more than two decades, and had studied French, in which gender differentiation is present. Thus, it is not readily apparent whether gender differentiation in a native language facilitated or obstructed the acquisition of an analogous language structure by participants in this study. What is clear is that repeated exposure to the stimuli and the contingencies of reinforcement was critical for correct performance in tests and that seems congruent with the behavior analytic view that language learning is a long and interactive process (Dale, 2004).

Physical differences in the endings of the singular and plural nouns permitted discrimination between members of the singular and plural classes. No additional training was required for participants to produce conditional relations among the new plural stimuli that paralleled the relations among singular stimuli that had been established in Experiment 1. This emergent behavior likely had its basis in the physical similarities of the singular and plural nouns and pictures, suggesting an important role for primary stimulus generalization in the emergence of the new plural stimulus classes. Results of this experiment also showed that when contextual control was established with one member of an equivalence class (i.e., a printed noun), participants responded similarly when another member of the equivalence class (i.e., a picture) was encountered in the same context for the very first time. This may therefore be an experimental analogue of the process that underlies productions of appropriate verbal behavior in novel situations even though there is no apparent history of reinforcement for that behavior. Our outcomes support the notion that the stimulus equivalence paradigm may model processes involved in the development of generalized and flexible linguistic behavior both with and without direct training.

GENERAL DISCUSSION

In Experiment 1, 5 normally capable participants with no prior knowledge of the Icelandic language learned conditional relations among spoken Icelandic singular nouns and corresponding pictures, and among those same spoken singular nouns and corresponding printed nouns. Tests then demonstrated that the trained conditional relations had the properties of equivalence; corresponding singular stimuli were substitutable for one another on match-to-sample trials. Next, without further training, participants demonstrated conditional relations among the spoken plural forms of the singular nouns, the corresponding pictures, and the corresponding printed nouns, in all possible combinations. Those emergent performances appeared to be based on physical similarities between the plural stimuli and the members of the singular-noun equivalence classes, illustrating an interaction of primary stimulus generalization and stimulus equivalence. Performances on subsequent tests that required discriminating singular from plural stimuli confirmed that stimulus classes of plural nouns (in spoken, printed, and picture form) had developed without explicit training. In sum, after equivalence classes of singular stimuli were established, primary stimulus generalization resulted in entirely new classes of plural stimuli.

Experiment 2 showed that, after some conditional relations among the stimuli from Experiment 1 were brought under contextual stimulus control, all Experiment 1 stimuli were treated as members of either singular, plural, masculine, or feminine nouns depending on the contextual stimulus present. This contextually controlled classification was also demonstrated by all participants on a sorting test, though some participants required reviews of the contextual control training and retesting before they produced that outcome. The sorting test, unlike the match-to-sample trials, assessed contextually controlled classification with all the stimuli available to the participant concurrently, including stimuli that had never been presented with the contextual stimuli during training. Results of the Experiment 2 sorting tests therefore lend external validity to the outcomes of the computerized contextual control training and testing, and support the inference that the contextual stimuli exerted independent, higher-order conditional stimulus control over equivalence class membership (cf. Bush et al., 1989; Lynch & Green, 1991).

The experiments reported here extended previous research by demonstrating that stimulus equivalence procedures can be used to teach skills like sight-word reading in a new language, as well as contextually controlled classification of words in that language, quite efficiently without verbal instructions. (In fact, one participant commented that she ''would love to learn a foreign language in this manner''). The training provided to the participants in these experiments was roughly analogous to the expriences of young first-language learners of Icelandic, who probably are not taught explicitly that spoken words, pictures, and printed words can be substitutable for one another, or that nouns can belong to classes that are labelled singular, plural, feminine, and masculine, until they are in elementary school. Our participants were never told the English meanings of the contextual stimuli TALA and KYN; instead, the functions of those stimuli, as well as the underlying equivalence classes, were established through differential reinforcement involving four- and five-term contingencies (cf. Sidman, 1986). Young Icelandic children may learn to discriminate the features that make nouns members of masculine, feminine, singular, or plural

classes depending on context through similar contingencies. The differential reinforcement may be implicit (in the form of demonstrations, repetitions, expansions, extensions, continuation, or interruption of verbal interactions by adult listeners), or explicit (in the form of approval and direct corrections). Of course, unlike young children, our participants had long histories of language learning when they entered these experiments, albeit not with Icelandic or any related languages. The extent to which those histories affected the outcomes of our experiments is unknown. Systematic replications with individuals with limited language repertoires (e.g., young typically developing children, individuals with developmental disabilities or brain injuries) would be illuminating. Nevertheless, our results appear to confirm that stimulus equivalence and contextual stimulus control procedures can be used fruitfully to conduct experimental analyses of complex performances with naturally occurring linguistic stimuli.

On a related note, it could be argued that it would have been more efficient to simply tell our participants the rules for classifying the Icelandic nouns. The instructions that were provided in these experiments were minimal, and consisted primarily of ''touch'' or ''touch one'' when a sample or a comparison array was displayed on the computer screen. Most of the complex and interacting types of stimulus control demonstrated here were established by the reinforcement contingencies. Therefore, the procedures used in these experiments could potentially be used to develop semantic relations and contextually controlled linguistic classes with individuals for whom verbal instructions are ineffective or impractical.

REFERENCES

Adams, B. J., Fields, L., & Verhave, T. (1993). Formation of generalized equivalence classes. *The Psychological Record*, *43*, 553–566.

Barnes, D., & Keenan, M. (1993). A transfer of functions through derived arbitrary and nonarbitrary stimulus relations. *Journal of the Experimental Analysis of Behavior*, *59*, 61–81.

Berko, J. (1958). The child's learning of English morphology. *Word*, *14*, 150–177.

Bush, K. M., Sidman, M., & de Rose, T. (1989). Contextual control of emergent equivalence relations. *Journal of the Experimental Analysis of Behavior*, *51*, 29–45.

Catania, A. C. (1992). Learning (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (1992). Using stimulus equivalence procedures to teach name–face matching to adults with brain injuries. *Journal of Applied Behavior Analysis*, *25*, 461–475.

Dale, R. (2004). Cognitive and behavioral approaches to language acquisition: Conceptual and empirical intersections. *The Behavior Analyst Today*, *5*, 336–359.

Dale, R., Roche, J., & Duran, N. (2008). Language is complex. *International Journal of Psychology and Psychological Therapy*, *8*, 351–362.

Dube, W. V. (1991). Computer software for stimulus control research with Macintosh computers. *Experimental Analysis of Human Behavior Bulletin*, *9*, 28–30.

Dube, W. V., & McIlvane, W. J. (1996). Some implications of a stimulus control topography analysis for emergent behaviour and stimulus classes. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 197–218). Amsterdam: Elsevier Science B.V.

Fields, L., Reeve, K. F., Adams, B. J., Brown, J. L., & Verhave, T. (1997). Predicting the extension of equivalence classes from primary generalization gradients: The merger of equivalence classes and perceptual classes. *Journal of the Experimental Analysis of Behavior*, *68*, 67–91.

Fields, L., Reeve, K. F., Adams, B. J., & Verhave, T. (1991). Stimulus generalization and equivalence classes: A model for natural categories. *Journal of the Experimental Analysis of Behavior*, *55*, 305–312.

Galizio, M., Stewart, K. L., & Pilgrim, C. (2004). Tipicality effects in contingency-shaped generalized equivalence classes. *Journal of the Experimental Analysis of Behav*ior, *82*, 253–273.

Gatch, M. B., & Osborne, J. G. (1989). Transfer of contextual stimulus function via equivalence class development. *Journal of the Experimental Analysis of Behavior*, *51*, 369–378.

Green, G., Sigurdardóttir, Z. G., & Saunders, R. R. (1991). The role of instructions in the transfer of ordinal functions through equivalence classes. *Journal of the Experimental Analysis of Behavior*, *55*, 287–304.

Griffee, K., & Dougher, M. J. (2002). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior*, *78*, 433–447.

Gruendel, J. M. (1977). Referential extension in early language development. *Child Development*, *48*, 1567–1576.

Hayes, S., Kohlenberg, B. S., & Hayes, L. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior*, *56*, 119–137.

Houmanfar, R., Hayes, L. J., & Herbst, S. A. (2005). An analog study of first language dominance and interference over second language. *The Analysis of Verbal Behavior*, *21*, 75–98.

Kohlenberg, B. S., Hayes, S., & Hayes, L. (1991). The transfer of contextual control over equivalence classes through equivalence classes: A possible model of social stereotyping. *Journal of the Experimental Analysis of Behavior*, *56*, 505–518.

Lane, S. D., Clow, J. K., Innis, A., & Critchfield, T. (1998). Generalization of cross-modal stimulus equivalence classes: Operant processes as components in human category formation. *Journal of the Experimental Analysis of Behavior*, *70*, 267–279.

Lazar, R. M., & Kotlarchyk, B. J. (1986). Second-order control of sequence-class equivalences in children. *Behavioral Processes*, *13*, 205–215.

Lowe, C. F., Horne, P. J., Harris, F. D. A., Randle, V. R. L. (2002). Naming and categorization in young children: vocal tact training. *Journal of the Experimental Analysis of Behavior*, *78*, 527–549.

Lynch, D. C., & Green, G. (1991). Development and crossmodal transfer of contextual control of emergent stimulus relations.

Journal of the Experimental Analysis of Behavior, *56*, 139–154.

Mackay, H. A. (1991). Stimulus equivalence: Implications for the development of adaptive behavior. In R. Remington (Ed.), *The challenge of severe mental handicap: An applied behaviour analytic approach* (pp. 235–259). London, England: John Wiley & Sons.

Mackay, H. A., & Fields, L. (2009). Syntax, grammatical transformation, and productivity: A synthesis of stimulus sequences, equivalence classes and contextual control. In R. A. Rehfeldt & Y. Barnes-Holmes (Eds.), *Derived relational responding: Applications for learners with autism and other developmental disabilities*: *A progressive guide for change*. Oakland, CA: New Harbinger.

Mackay, H. A., Wilkinson, K. M., Farrell, C., & Serna, R. W. (2011). Evaluating merger and intersection of equivalence classes with one member in common. *Journal of the Experimental Analysis of Behavior*, *96*, 87–105.

McIlvane, W. J., Dube, W. V., Green, G., & Serna, R. W. (1993). Programming conceptual and communication skill development: A methodological stimulus class analysis. In A. P. Kaiser & D. B. Gray (Eds.), *Enhancing children's communication: Research foundations for intervention*. (pp. 243–285). Baltimore: Paul H. Brookes.

Pérez-González, L. A., & Serna, R. W. (2003). An analysis of generalized contextual control of conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *79*, 395–408.

Pilgrim, C., & Galizio, M. (1996). A class of correlations or a correlation of classes. In T. R. Zentall, & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 173–197). Amsterdam: Elsevier.

Salzinger, K. (2008). Skinner's *Verbal Behavior*. *Journal of Psychology and Psychological Therapy*, *8*, 287–294.

Saunders, R. R., Saunders, K. J., Kirby, K. C., & Spradlin, J. E. (1988). The merger and development of equivalence classes by unreinforced conditional selection of comparison stimuli. *Journal of the Experimental Analysis of Behavior*, *50*, 145–162.

Saunders, R. R., Wachter, J., & Spradlin, J. E. (1988). Establishing auditory stimulus control over an eight-member equivalence class via conditional discrimination procedures. *Journal of the Experimental Analysis of Behavior*, *49*, 95–115.

Serna, R. W., & Pérez-González, L. A. (2003). Transfer of specific contextual functions to novel conditional discriminations. *Journal of the Experimental Analysis of Behavior*, *79*, 383–393.

Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, *14*, 5–13.

Sidman, M. (1986). Functional analysis of emergent verbal classes. In T. Thompson & M. D. Zeiler (Eds.), *Analysis and integration of behavioral units* (pp. 213–245). Hillsdale, NJ: Erlbaum.

Sidman, M. (1994). *Equivalence relations and behavior: A research story*. Boston: Authors Cooperative.

Sidman, M., Rauzin, R., Lazar, R., Cunningham, S., Tailby, W., & Carrigan, P. (1982). A search for symmetry in the conditional discriminations of rhesus monkeys, baboons, and children. *Journal of the Experimental Analysis of Behavior*, *37*, 23–44.

Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, *37*, 5–22.

Sidman, M., Wynne, C. K., Maguire, R. W., & Barnes, T. (1989). Functional classes and equivalence relations. *Journal of the Experimental Analysis of Behavior*, *52*, 261–274.

Sigurðardóttir, Z. G., Green, G., & Saunders, R. R. (1990). Equivalence classes generated by sequence training. *Journal of the Experimental Analysis of Behavior*, *53*, 47–63.

Skinner, B. F. (1957). *Verbal behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Stemmer, N. (1990). Skinner's *Verbal Behavior*, Chomsky's review, and mentalism. *Journal of the Experimental Analysis of Behavior*, *54*, 307–315.

Thomson, J. R., & Chapman, R. S. (1977). Who is ''Daddy'' revisited: the status of two-year-olds' over-extended words in use and comprehension. *Journal of Child Language*, *4*, 359–375.

Washio, Y., & Houmanfar, R. (2007). Role of contextual control in second language performance. *The Analysis of Verbal Behavior*, *23*, 41–56.

Appendix 1
Experiment 1 trial types. Trials were presented in unsystematic order within sessions.

Stimulus set	Relations trained	Sample	Comparisons		
TRAIN					
1	A1B1	A1	B1	B2	B3
	A2B2	A2	B3	B1	B2
	A3B3	A3	B2	B3	B1
	A4B4	A4	B4	B5	B6
	A5B5	A5	B6	B4	B5
	A6B6	A6	B5	B6	B4
2	A1C1	A1	C1	C2	C3
	A2C2	A2	C3	C1	C2
	A3C3	A3	C2	C3	C1
	A4C4	A4	C4	C5	C6
	A5C5	A5	C6	C4	C5
	A6C6	A6	C5	C6	C4
TEST EQUIVALENCE					
1	C1B1	C1	B1	B2	B3
	C2B2	C2	B3	B1	B2
	C3B3	C3	B2	B3	B1
	C4B4	C4	B4	B5	B6
	C5B5	C5	B6	B4	B5
	C6B6	C6	B5	B6	B4
2	B1C1	B1	C1	C2	C3
	B2C2	B2	C3	C1	C2
	B3C3	B3	C2	C3	C1
	B4C4	B4	C4	C5	C6
	B5C5	B5	C6	C4	C5
	B6C6	B6	C5	C6	C4
TEST PRIMARY STIMULUS GENERALIZATION					
1	G1F1	G1	F1	F2	F3
	G2F2	G2	F3	F1	F2
	G3F3	G3	F2	F3	F1
	G4F4	G4	F4	F5	F6
	G5F5	G5	F6	F4	F5
	G6F6	G6	F5	F6	F4
	F1G1	F1	G1	G2	G3
	F2G2	F2	G3	G1	G2
	F3G3	F3	G2	G3	G1
	F4G4	F4	G4	G5	G6
	F5G5	F5	G6	G4	G5
	F6G6	F6	G5	G6	G4
2	E1G1	E1	G1	G2	G3
	E2G2	E2	G3	G1	G2
	E3G3	E3	G2	G3	G1
	E4G4	E4	G4	G5	G6
	E5G5	E5	G6	G4	G5
	E6G6	E6	G5	G6	G4
	E1F1	E1	F1	F2	F3

Appendix 1, *cont.*

Stimulus set	Relations trained	Sample		Comparisons	
	E2F2	E2	F3	F1	F2
	E3F3	E3	F2	F3	F1
	E4F4	E4	F4	F5	F6
	E5F5	E5	F6	F4	F5
	E6F6	E6	F5	F6	F4

Appendix 2

Experiment 2 contextual control training trial types. In each trial block, each positive comparison (S+) appeared equally often on the left and right.

Relations trained	Contextual Stimulus	Sample	S+	S−
TALA	O1P1	TALA	O1	P1	R1/R2/R3
	O1P2			P2	
	O1P3			P3	
KYN	O1R1	KYN	O1	R1	P1/P2/P3
	O1R2			R2	
	O1R3			R3	
TALA	Q1R1	TALA	Q1	R1	P1/P2/P3
	Q1R2			R2	
	Q1R3			R3	
KYN	Q1P1	KYN	Q1	P1	R1/R2/R3
	Q1P2			P2	
	Q1P3			P3	
TALA	O2P1	TALA	O2	P1	R1/R2/R3
	O2P2			P2	
	O2P3			P3	
KYN	O2R1	KYN	O2	R1	P1/P2/P3
	O2R2			R2	
	O2R3			R3	
TALA	Q2R1	TALA	Q2	R1	P1/P2/P3
	Q2R3			R2	
	Q2R3			R3	
KYN	Q2P1	KYN	Q2	P1	R1/R2/R3
	Q2P2			P2	
	Q2P3			P3	
TALA	O3P1	TALA	O3	P1	R1/R2/R3
	O3P2			P2	
	O3P3			P3	
KYN	O3R1	KYN	O3	R1	P1/P2/P3
	O3R2			R2	
	O3R3			R3	
TALA	Q3R1	TALA	Q3	R1	P1/P2/P3
	Q3R2			R2	
	Q3R3			R3	
KYN	Q3P1	KYN	Q3	P1	R1/R2/R3
	Q3P2			P2	
	Q3P3			P3	

Appendix 3
Singular vs. Plural discrimination test trials

Test set	Relation tested	Sample	Comparisons		
1	C1B1	C1	B1	F1	B3
			F1	B1	F3
			B5	F1	B1
			F5	B1	F1
			B1	B3	F1
			F1	F5	B1
2	G1F1	G1	B1	F1	B3
			F1	B1	F3
			B5	F1	B1
			F5	B1	F1
			B1	B3	F1
			F1	F5	B1
3	C3B3	C3	B3	F3	F1
			F3	B3	B1
			F5	F3	B3
			B5	B3	F3
			B3	F1	F3
			F3	B5	B3
4	G3F3	G3	B3	F3	F1
			F3	B3	B1
			F5	F3	B3
			B5	B3	F3
			B3	F1	F3
			F3	B5	B3
5	C5B5	C5	B5	F5	B1
			B5	F1	F5
			F5	B5	B3
			F5	B3	B5
			B1	B5	F5
			F3	F5	B5
6	G5F5	G5	B5	F5	B1
			B5	F1	F5
			F5	B5	B3
			F5	B3	B5
			B1	B5	F5
			F3	F5	B5
7	C2B2	C2	B2	F2	B4
			F2	B2	F4
			B6	F2	B2
			F6	B2	F2
			B2	B4	F2
			F2	F6	B2
8	G2F2	G2	B2	F2	B4
			F2	B2	F4
			B6	F2	B2
			F6	B2	F2
			B2	B4	F2
			F2	F6	B2
9	C4B4	C4	B4	F4	F2

Appendix 3, *cont.*

Test set	Relation tested	Sample	Comparisons		
			F4	B4	B2
			F6	F4	B4
			B6	B4	F4
			B4	F2	F4
			F4	B6	B3
10	G4F4	G4	B4	F4	F2
			F4	B4	B2
			F6	F4	B4
			B6	B4	F4
			B4	F2	F4
			F4	B6	B3
			F3	B5	B3
11	C6B6	C6	B6	F6	B2
			B6	F2	F6
			F6	B6	B4
			F6	B4	B6
			B2	B6	F6
			F4	F6	B6
12	G6F6	G6	B6	F6	B2
			B6	F2	F6
			F6	B6	B4
			F6	B4	B6
			B2	B6	F6
			F4	F6	B6

Appendix 4

Experiment 2 contextual control test trial types. In each trial block, each positive comparison (S+) appeared equally often on the left and right.

Relations tested		Contextual Stimulus	Sample	S+	S−
TALA	K1L1	TALA	K1	L1	N1/N2/N3
	K1L2			L2	
	K1L3			L3	
KYN	K1N1	KYN	K1	N1	L1/L2/L3
	K1N2			N2	
	K1N3			N3	
TALA	M1N1	TALA	M1	N1	L1/L2/L3
	M1N2			N2	
	M1N3			N3	
KYN	M1L1	KYN	M1	L1	N1/N2/N3
	M1L2			L2	
	M1L3			L3	
TALA	K2L1	TALA	K2	L1	N1/N2/N3
	K2L2			L2	
	K2L3			L3	
KYN	K2N1	KYN	K2	N1	L1/L2/L3

Appendix 4, *cont.*

Relations tested	Contextual Stimulus	Sample	S+	S−
	K2N2		N2	
	K2N3		N3	
TALA	M2N1	TALA	M2	N1	L1/L2/L3
	M2N2		N2	
	M2N3		N3	
KYN	M2L1	KYN	M2	L1	N1/N2/N3
	M2L2		L2	
	M2L3		L3	
TALA	K3L1	TALA	K3	L1	N1/N2/N3
	K3L2		L2	
	K3L3		L3	
KYN	K3N1	KYN	K3	N1	L1/L2/L3
	K3N2		N2	
	K3N3		N3	
TALA	M3N1	TALA	M3	N1	L1/L2/L3
	M3N2		N2	
	M3N3		N3	
KYN	M3L1	KYN	M3	L1	N1/N2/N3
	M3L2		L2	
	M3L3		L3	

Appendix 5
Age and gender of participants, gender differentiation in native language and other languages spoken

Participant	Age	Gender	Gender differentiation in native language	Gender differentiation in other languages spoken
M.B.*	54	F	No (Armenian)	N/A
R.C.	35	F	No (English)	N/A
T.M.	26	F	No (Malayalam)	No (English) Yes (French)
M.T.	30	M	Yes (Arabic)	No (English)
C.S.*	15	F	Yes (Portuguese)	N/A

* was a novice in English